

Bio-Inspired Probabilistic Model for Crowd Emotion Detection

Mirza Waqar Baig, Emilia I. Barakova, Lucio Marcenaro, Carlo S. Regazzoni
and Matthias Rauterberg

Abstract—Detection of emotions of a crowd is a new research area, which has never, to our knowledge, been accounted for research in previous literature. A bio-inspired model for representation of emotional patterns in crowds has been demonstrated. Emotions have been defined as evolving patterns as part of a dynamic pattern of events. This model has been developed to detect emotions of a crowd based on the knowledge from a learned context, psychology and experience of people in crowd management. The emotions of multiple people making a crowd in any surveillance environment are estimated by sensors signals such as a camera and are being tracked and their behavior is modeled using bio-inspired dynamic model. The behavior changes correspond to changes in emotions. The proposed algorithm involves the probabilistic signal processing modelling techniques for analysis of different types of behavior, interaction detection and estimation of emotions. The emotions are recognized by the expectation of temporal occurrences of causal events modeled by Gaussian mixture model. The model has been evaluated using the simulated behavioral model of a crowd.

I. INTRODUCTION

Emotion is a feeling evoked by environmental stimuli or by internal body states [1], which modulates human behavior in terms of actions in the environment or changes in the internal states of the body. The idea of emotional expressions was first recognized by Darwin, who discovered that the natural selection process is not only applicable to anatomic patterns but also can be applied to animal behavior and mind.

Emotion is a vital part of human behavior, which is derived from a Latin word which means “move”. It depends on the influences of the external factors such as interaction with people and environmental conditions. In interactive cognitive environments, the evolution of smart spaces is defined by the ambient intelligence paradigms. Emotions take

their shape according to the conditions of the environment around and take part in response according to that situation.

Emotions are brain processes which evolve to regulate behaviors under different encountered situations. The decision making about these different encountered situations is governed by the evaluation dimension added by emotions, which plays a vital role in it. Emotions are produced as a result of complex (conscious or unconscious) cognitive processes [2]. The components of emotions can be defined as affect, cognitive reaction, physiological reaction and behavioral reaction. Affect is a subjective feeling such as happiness, sadness, anger, etc. The affect can be taken to indicate an instinctive reaction for simulation occurring before the cognitive processes. In cognitive reaction, you recognize, or know what happened. The physiological changes (autonomic arousal) bring the internal changes that alter the autonomic nervous system and hormones. The behavioral reaction (non-verbal) is the feeling disposed to behave in particular ways, depending on the emotion (e.g., cursing someone when angry). The effects of the complex emotional cognitive process (detection/arousal/response) are the actions of the cognitive entity as it interacts with the external world or possibly with other entities as well as state changes of the cognitive entity externally visible to the interacting entity. The idea of recognizing and modeling of emotions for interactions with situations can be designed on the basis of intelligent cognitive environment systems, which can be defined by the evolution of smart spaces defined by the ambient intelligent paradigms [3]. This involves the two parts of context awareness and interaction with users. The context awareness includes the probabilistic models for human behaviors, situation awareness and recognition of emotional content of human actions. Interactions with users involve the environmental control to influence behaviors, bio-inspired embodied cognition for decision and threat management.

The detection of emotions of crowd is an emerging research area and new developments have been made recently. Previously, research has been carried out in single person emotion detection and classification. For instance, facial emotion expression detection systems have been utilized to detect distinctive facial features and calibrate them into emotions [4] [5] of individuals. Moreover, the Facial Action Coding System (FACS) [6] developed by Ekman and Friesen uses a set of action units (AUs) that model emotions on independent motions of the face. Voice expressions have also been used to detect emotions of individuals. The audio-based affect recognition using the features of pitch and context has been developed such as those in [7] [8]. Furthermore, the understanding of frequency and pitch of the words in speech and understanding their meaning with

Mirza Waqar Baig is an EMJD ICE fellow at both Eindhoven University of Technology, P.O. Box 513, 5600 MB, Netherlands and University of Genoa, Via Opera Pia 11a, Genoa 16145, Italy (phone: +31-6-15507045; e-mail: M.W.Baig@tue.nl).

Emilia I. Barakova was with RIKEN Brain Institute Japan. She is now with the Department of Industrial Design as assistant professor, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Netherlands (e-mail: E.I.Barakova@tue.nl).

Lucio Marcenaro is assistant professor at Department of Biophysical and Electronic Engineering, University of Genoa, 16145 Genoa, Italy (e-mail: Lucio.Marcenaro@unige.it)

Carlo S. Regazzoni is professor at Department of Biophysical and Electronic Engineering, University of Genoa, 16145 Genoa, Italy (e-mail: Carlo@dibe.unige.it)

Matthias Rauterberg is professor at Industrial Design Department, Eindhoven University of Technology, P.O. Box 513, 5600 MB, Netherlands (e-mail: G.W.M.Rauterberg@tue.nl).

culture and situation context have been useful in emotion detection studies as shown in [9] [10]. Additionally, body language and posture are useful features in human emotion detection. Specifically, the predefined body language patterns and postures are guessed by the activity recognizers. The transition of body patterns over a specific interval of time, posture patterns to describe the interest towards some specific things are quite useful in detecting emotions [11][12][13][14]. Physiology based affect detection is quite a broad approach, which requires the inclusion of delicate and calibrated sensors on body [15] and in some cases quite strict supervision, as in case of brain imaging and electroencephalography (EEG) signals [16][17]. The physiology based methods are quite accurate but they are usually used to calibrate the ground truth for emotion and affect detection. They cannot be used in real life situations. Apart from these techniques, there are also multimodal techniques, which fuse many features such as face, voice, and posture etc. Specifically, the data coming from different sensors is fused to make a decision of emotion [18][19]. All the compact review mentioned above is in use for individual emotion detection, these systems are not being employed in crowd emotion detection.

The emotion of the crowd detection is of high importance. There has been a plethora of studies on crowd on crowd behavior management but with emotion is one step before behavior. You can guess the future situation if you know how people are feeling i.e., emotions and can avoid it by implementing some pre-designed safety plans. Consider a football field which is under being surveillance by cameras and the visual feedback is composed by several monitors. The behavioral change can be modeled when it has started occurring, so reaction time is reduced. In fact, if emotions are being tracked in the crowd. Then we can make better predictions of what will happen. The main shortcoming of detection of emotions in a crowd is that we have to solely depend on video cameras; we cannot use delicate features such as face emotion features [6], activity features [12] and voice features [8]. The processing speed and viewpoint differences are main constraints in computation in crowds and this worsens during abnormal situations. Also, face recognition slows down when the number of people increase, the voice gets mixed up among many participants with in a crowd scenario. For example, in a football stadium confusion ensues when the posture and body features require the body and posture to be visible and calibrated together and it is impossible to make a network of sensors to record body and posture features within a crowd. Therefore, the features available for crowd emotion detection are quite crude, but using the knowledge from psychology and learning the context and history, we are able to develop a model that can detect the emotions of the crowd. Gustave Le Bon [20] described. A crowd acts like a collective mind and acts and reacts. Consequently, this phenomenon becomes more common when some abnormal situation arises. Therefore, our evaluations mostly correspond to the accuracy due to features

we have selected.

The main contribution of this paper is to develop the first model to detect the emotions of the crowd. Moreover, this paper presents the dynamic Bayesian probabilistic modelling of autobiographical memory based on modelling of events on the basis of psychological theories and development of a realistic behavioral simulation model for simulating crowd behavior. The evaluations have been performed for training and testing purposes.

The remaining work of this paper is as follows: Section II of paper describes the proposed research and in Section III, the experiments and results have been discussed. Section IV presents the conclusion.

II. PROPOSED FRAMEWORK

The proposed framework is based on the collection of the local features from a crowd and using the Damasio [21] findings to make these observations in causal relationships based on dynamic Bayesian networks. These causal relationships are then being clustered using SOM [22] to define events. The Ortony, Clore and Collins (OCC) theory [23] are used to define causes of emotions for learning and classification and for calibrating the model. Russell et al. circumplex model of affect [24] is being used for calibrating the emotion on valance scale.

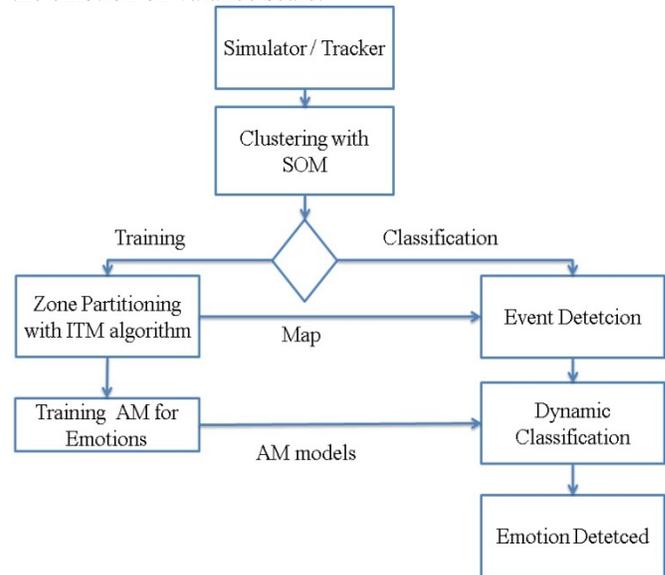


Figure 1: Block Diagram of Proposed Algorithm

The Figure 1 shows the flowchart of proposed algorithm. The emotion modelling and classification in a crowd is based on the OCC theories [16], in which emotions are defined as valenced reaction to the events, agents and objects. The developed theoretical approach of OCC is the assumption that emotions develop as a consequence of certain cognitive processes and their interpretations. Therefore, OCC model involves the cognitive triggering of emotions due to events, agents and objects. Each cognitive trigger has its causes due to some local variable such as urgency for an event or for other liking or dis-liking of an event, liking or dis-liking of an agent's action and the attractiveness for an object. Each

emotion evolves due to unique fusion of each of these variables, which constitutes the cognitive structure of such an emotion. These elicitations in our model are being modeled by the local variables from sensors. The interaction association among people is modelled on the basis of neurophysiologist A. Damasio [21]. Damasio explains the self-conscious is influenced by the conscious interactions with the environment. The brain manipulates the internal state of humans and its relationship with external factors that includes agents, objects and events, defined as proto self and core self. These relationships are stored as visual interpretations the brain's relationships with objects and feelings emerged due to each kind of interaction with the objects. These are stored in the form of images and these images are sensory perceptions of brain interactions. These imagery interaction memories are dependent on visual sensory inputs as well as other sensors such as auditory, somatosensory etc. The images are managed by a pattern of neural networks; the neural network involved in interpretations of interactions and activities constitutes the neural patterns with the help of sensory inputs. The proto self and core self, which are the visual interpretations/representations constituted by neural network patterns can be represented as first order neural networks. The causal interaction patterns is between organism and object generate core conscious. Moreover, second order neural patterns that are generated due to the causal relationships among different entities and its effects from inputs from visual perception representations, which modifies itself into proto self. Damasio also claimed that the core conscious is continuously evolving due to interaction with external perceptions.

The neurophysiological observations mentioned above are used to model interactions among the people. The interactions are causal relationships inspired from the Damasio theory. The causal relationships among the people are the basis of the features of our model.

A. Topological Partitioning

In order to extract the most relevant information from the crowd environment, it needs to be mapped into virtual sub-parts. These subparts are known as zones. A regular i.e. grid type partitioning of 2-D environment can be used, but it will not take into the account, the specific contextual information about the environmental elements such as walls, gates or obstacles of unknown shape in the environment. This map discretised the 2-D space into the areas where the observations appear. The spatial information can be extracted from the generated map using dimensionality reduction algorithms. This process does not only reduce the data dimensions but also the spatial distribution of data distribution remains unaltered. The topology representing networks (TRN) [25] algorithms are effective in defining the observation area into dynamic or definite areas using vector quantization and then creating the similarity based links among the centres of neighbouring areas. One of these algorithms, instantaneous topological map (ITM) [26] algorithm, is employed to partition the environment into

different zones. It was proposed by Jockusch and Ritter in 1999, and they demonstrated that ITM is a suitable approach to form discretised zones of temporal correlated trajectories. Given an observation of $X_n(t)$ position of an object (agent) A_n at time t , where $X_n(t) \in \mathbb{R}^2$. A trajectory is a sequence of $X_n = X_{n_{t_1}}, \dots, X_{n_{t_i}}$ where t_1 is the first observation instant and t_k is the last observation instant. A set of trajectories have been obtained and repeated over different scenarios to increase the trajectory deviations. These trajectories $\{X_n\}_{n=1}^N$, where N is the total number of trajectories observations are obtained, used to create discretised mapping of observation space. The ITM map $\psi \subseteq \mathbb{R}^2$, then the ITM algorithm produces a transformed set of given observations $\{b_k\}_{k=1}^k$ such that $b_k \cup b_i = \emptyset$ for $k \neq i$ and $\cup_{k=1}^k b_k = B \subseteq \psi$. The ITM zones are a subset $\{c_k\}_{j=1}^j$ from $\{b_k\}_{k=1}^k$ of maximum information capacity for zones such that $c_j \cup c_h = \emptyset$ for $h \neq i$ and $\cup_{j=1}^j c_j = C \subseteq B$. After iterating through several instances, the zone merging b_k i.e. $i = 1, \dots, l \neq k, b_k \subseteq b_i$. It produces the map with c_k zones, which have the definite centre and the length of zone is defined according to the instance of the zone and zone width is analogous to insertion parameter threshold τ as shown in Figure 2. The τ defines the minimum distance between the given zone c_k given all the observation set b_k for creating new zone.

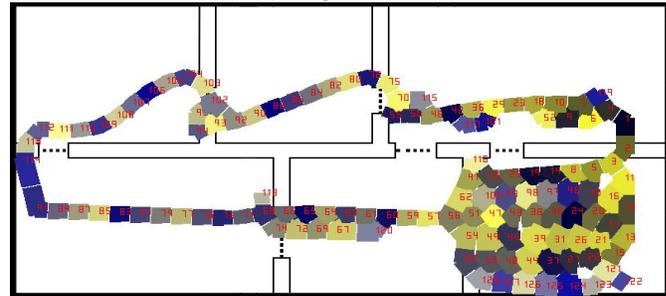


Figure 2: $\tau = 500$, τ defines the minimum distance between two consecutive zones for creating new zone given all the observations

B. Probabilistic Framework

The ability to predict the emotions requires predicting actions before they happen in the area under observation and surveillance in order to define the correct interactions and activities that lead to emotions. To reach this goal, a possible technique is to memorize all possible interactions that occur in a related area among people and the environment and their reactions due to these actions. A bio-inspired approach is used to encode the interactions that has been developed based on the work described in [27][28][29][30]. These interactions are generated due to cause and effect relationships that happen between agents, agents and objects and agents and environment. The interactions can be of two types 1) the effects of interacting entities on an entity is examined by how it changes its internal state of the entity (passive interaction) 2) the effects on external entities as a reaction due to the

action of the entity is examined by how it affects the interacting entities(active interaction). These types of interactions can be learned using autobiographical memory as mentioned by Damasio. To describe the effect of interacting quantities on internal and external states, the proto (internal) and core (mirror of external) states are defined. The proto and core observations are defined as proto (entity) and core (crowd). It makes a chain of temporally and spatially aligned proto \vec{x}_p and core state \vec{x}_c vectors. The proto and core events can be defined as ε_p and ε_c using the probabilistic model to develop autobiographical memory (AM). The triplets of events for passive and active interactions are $\{\varepsilon_p^-, \varepsilon_c, \varepsilon_p^+\}$ and $\{\varepsilon_c^-, \varepsilon_p, \varepsilon_c^+\}$, respectively. They represent the causal relationship in terms of initial situation (first event $\varepsilon_{p,c}^-$), the cause (second event $\varepsilon_{c,p}$), and the consequent effect of the examined entity (third event $\varepsilon_{p,c}^+$). The sequence of these states can be decided by using a probabilistic graphical model that describes the relationships among them using statistical and mathematical similarities among interactions. The interactions consist of temporal sequences of interdependent states that give rise to stochastic process described by two probability distributions: $p(\varepsilon_p^+ | \varepsilon_c, \varepsilon_p^-)$ and $p(\varepsilon_c^+ | \varepsilon_p, \varepsilon_c^-)$. The Dynamic Bayesian networks (DBN) are used to model the interaction as shown in Figure 3. The conditional probability densities (CPD) $p(\varepsilon_t^p | \varepsilon_{t-1}^p)$ and $p(\varepsilon_t^c | \varepsilon_{t-1}^c)$ simulates the motion tracking and crowdedness patterns. The interactions between the two interacting objects can be modelled using conditional probability densities (CPDs):

$$p(\varepsilon_t^p | \varepsilon_{t-\Delta t}^c) \quad (1)$$

$$p(\varepsilon_t^c | \varepsilon_{t-\Delta t}^p) \quad (2)$$

The (1) represents the probability that events ε_c occurred at time $t - \Delta t^c$, by the interacting object, which is related to the core context and vice versa in case of (2). The casual relationships between the two interacting objects are modelled using two conditional probabilities (CPDs):

$$p(\varepsilon_t^p | \varepsilon_{t-\Delta t}^c, \varepsilon_{t-\Delta t}^p) \quad (3)$$

$$p(\varepsilon_t^c | \varepsilon_{t-\Delta t}^p, \varepsilon_{t-\Delta t}^c) \quad (4)$$

The probability densities in (3) and (4) consider interaction of (1) and (2) as well as initial situation $\varepsilon_{t-\Delta t}^c$ and $\varepsilon_{t-\Delta t}^p$. The observations associated with proto and core is:

$$\varepsilon_t^p = v(x, y)_t \quad (5)$$

where $(x, y)_t = \frac{d(x, y)_t - d(x, y)_{t-1}}{\Delta t}$, $v(x, y)_t$ is the velocity of entity under observation at time t and $d(x, y)_t$ is the position of entity at time t .

The core observation is given by:

$$\varepsilon_t^c = d_t \quad (6)$$

where d_t of entities in a given area. Therefore, the proto and core states are:

$$\vec{x}_p(t) = \{x_{p1}(t), \} \quad (7)$$

In order to represent these elements with respect to contextual information, a clustering technique is employed. To retain the large data dimensionality with main representations of vectors, many algorithms have been described for dimensionality reduction. A self-organized map

(SOM), unsupervised classifier is used to convert multidimensional proto $\vec{x}_p(t)$ and core vectors $\vec{x}_c(t)$ into low dimension W-Z, where W is the dimension of map (layer) in which each cluster consists of clusterization of related vectors according to their homogeneity. The clusterization process, maps proto and core states into 2-D vectors, which correspond to the neurons of the SOM map. They are called core super-states S_{x_c} and proto super-states S_{x_p} .

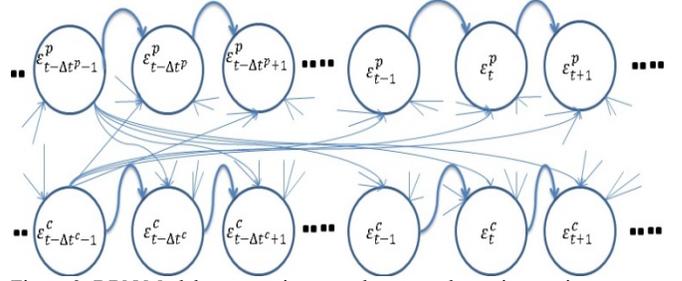


Figure 3: DBN Model representing causal proto and core interactions

The choice of SOM for dimensionality reduction and clustering is mainly due to its capability to retain the major information from features in a plausible mathematical way the clustering behaviour is based on winner takes-all demonstrated by distributed bio-inspired decision mechanisms. SOM map layer formed by M neurons, dimensions are used in order to find best matching unit (BMU) (S, w) such that $S \times w = M$. The number of proto and core states will be then M and possible core and proto events will be M^2 . SOM allows clustering the proto $\vec{x}_p(t)$ and core vectors $\vec{x}_c(t)$ into corresponding neuron map as super-states, which are then known as proto super-state S_{x_p} and core super-state S_{x_c} . The parameter M is to be tuned. The labels associated are given by:

$$S_{x_p}^i \mapsto l_p^i, \quad i = 1, \dots, n_p \quad (8)$$

$$S_{x_c}^j \mapsto l_c^j, \quad j = 1, \dots, n_c \quad (9)$$

Where $S_{x_p}^i$ and $S_{x_c}^j$ are the i^{th} and j^{th} super-states, and n_p and n_c are total number of proto and core states generated during mapping.

An event is defined by the sequential changes in proto and core states. The proto and core super-states are related with the semantic labels from the ITM zones. An event is defined when a proto or core super-state changes its zone. It gives rise to autobiographical memory associated with each zone. Using the SOM representation, it is possible to detect changes in the map through super-states, where super-states are connected by local features for particular instances. This representation encompasses the changes in state vectors $\vec{x}_p(t)$ and $\vec{x}_c(t)$ for every time instant as movements in map. If changes in state vectors $\vec{x}_p(t)$ and $\vec{x}_c(t)$ do not imply changes in super-state labels $S_{x_p}^i \mapsto l_p^i$ and $S_{x_c}^j \mapsto l_c^j$, then the SOM mapping needs to be recalibrated as the semantics defined are not the correct representation of events. When super-states $S_{x_p}^i$ and $S_{x_c}^j$ change during specific time instants, it contextual modification entails an event. Therefore, an event is defined as: $\theta_e^t = l_{p,c}^{i,j}(t-1) \mapsto l_{p,c}^{i,j}(t)$ where $i, j = 1, \dots, n_{p,c}$, with timing constraints T_{max} . There are also null events (null

changes in super-states $i = j$ can be defined as $\theta_e^t = \emptyset_e$. It gives rise to autobiographical memory model, which is modelled by the events in which learning the changes from proto super-states to core-super-states and subsequent modification of core super-state is memorized. Considering the core event θ_c occurring at T_1 , the changes due to core event θ_c on the internal states must be tracked in order to describe how the external interaction takes place, due to which a core event is triggered and core event is generated, occurred. To track this causal relationship, a temporal assumption of timing should be adopted by defining a temporal window with duration T_{max}^- to detect what the proto super-state $S_{x_p}^-(t)$, with timing window $T_1 - T_{max}^- < t < T_1$ (i.e., the initial state before core event was provoked) and its succeeding alteration due to core event, i.e., $S_{x_p}^+(t)$ with $T_1 < t < T_1 + T_{max}^+$, where T_{max}^+ is the relaxation time after which the core event has done proto modification. The following three types of events can be memorized.

- 1) $\theta_p^- = S_{x_p}^- \rightarrow S_{x_p}^-$: is the proto event at initial time instant. This represents the alteration of the proto super-state from $S_{x_p}^0 \mapsto l_p^i$ to $S_{x_p}^- \mapsto l_p^j$ that can happen before the core event. The event θ_p^- also remembers the time window T_{max}^- . The two labels l_p^i and l_p^j are related with super-states $S_{x_p}^0$ and $S_{x_p}^-$. The event θ_p^- also remembers the time window T_{max}^- .
- 2) $\theta_c = S_{x_c}^- \rightarrow S_{x_c}^+$: is the core event. It shows the alteration of external super-state from $S_{x_c}^- \mapsto l_c^m$ to $S_{x_c}^+ \mapsto l_c^n$.
- 3) $\theta_p^+ = S_{x_p}^- \rightarrow S_{x_p}^+$: is the proto event following the core event. It is also associated with the super-states as $S_{x_p}^- \mapsto l_p^j$ to $S_{x_p}^+ \mapsto l_p^k$.

The following triplet $\{\theta_p^-, \theta_c, \theta_p^+\}$ represents the self-abstraction, which is related with autobiographical memory. The autobiographical memory represents the core conscious in Damasio work. The autobiographical memory model of interaction has following assumptions: 1) The events triplets will be stored as passive memory (proto-core-proto); 2) One core event initiates and produces one proto super-state; 3) The event only happens if proto event changes following the core event during duration of T_{max}^+ ; and 4) If proto event preceding the core event, it must happen during T_{max}^- .

This memory now consists of different instances of a simulated environment. The offline learning of autobiographical memory involves the difference between positive emotions and negative emotions instantiated into AM coupled events. According to OCC theory, the emotions changes in reaction to events, agents and objects. We use AM events, which are already coupled by probability density of observation from agents and events. To obtain the events relevant to negative emotions, there is no ground truth available to distinguish between positive and negative emotions. To scale the emotion using Russell et. al [24], a two dimensional map of mental space of emotions has been used.

On this map, a valance scale to define emotions into positive and negative emotions. To define the negative emotions, we used collected the data of different type of abnormal situations in which the people have negative emotions, this collected data was also discussed with experts in this field. We define the negative emotions on the basis of famous situations such as (i) Herding behaviour [31] (ii) Turbulent flow [32] (iii) Stop-and-go waves [33]. The learning is done by taking coupled proto events θ_p and core events θ_c . The training sequence is composed of vector:

$$E_{k_w} = \{E_{1_w}^p; E_{1_w}^c; E_{2_w}^p; E_{2_w}^c, \dots, E_{N_w}^p; E_{N_w}^c\} \quad (10)$$

where w is the tag of vector belonging to normal or abnormal event, which in turn specifies to positive and negative emotions. Every time three consecutive events $\{\theta_{t-\Delta t}^{p(i,j)}, \theta_{t-\Delta t}^{p(m,n)}, \theta_t^{p(j,k)}\}_w$ are observed, the frequency of occurrence of this event is increased each time by a vote. Each time the estimation of occurrence of event $\theta^{p(j,k)}$, preceded by two events $\theta^{c(m,n)}$ and $\theta^{p(i,j)}$, $p(\theta^{p(j,k)} | \theta^{c(m,n)}, \theta^{p(i,j)})$ of estimation is obtained by normalizing the frequency of event over the last event $\theta^{p(i,j)}$. The temporal evolution of events is also memorized in order to develop the CPD of events. The Gaussian mixture model GMM is used to encode the temporal evolution of events. The Figueiredo and Jain [34] GMM model is used.

$$p(\Delta t^p | \theta^{p(j,k)}, \theta^{c(m,n)}, \theta^{p(i,j)}) = \sum_{i=1}^{N_k} \pi_i \mathfrak{N}(\Delta t^p | \mu_i, \Sigma_i \theta_n) \quad (11)$$

$$p(\Delta t^c | \theta^{c(m,n)}, \theta^{p(j,k)}, \theta^{c(n,p)}) = \sum_{i=1}^{N_k} \pi_i \mathfrak{N}(\Delta t^c | \mu_i, \Sigma_i \theta_n) \quad (12)$$

where N_k is the number of modes in GMM..

C. Emotion Recognition

An accumulative measure of online emotion classification is proposed in the task. Whenever an observed external event is detected θ_{obs}^c , the proto map is analysed which was the internal event θ_w^p . The AM is then examined to describe the internal event belonging to which emotion class w , positive or negative. The probability distribution estimated by GMM of detected event and the one derived from the map is evaluated for matching by Hellinger Distance [35]. The Hellinger distance is evaluated

$$\Delta^2(\theta_{obs}^c, \theta_w^p) = 1 - \frac{\sqrt{2\sigma_1\sigma_2}}{\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{1}{4} \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}} \quad (13)$$

where $\theta_{obs}^c \sim \mathfrak{N}(\sigma_1, \mu_1)$ and $\theta_w^p \sim \mathfrak{N}(\sigma_2, \mu_2)$

III. EXPERIMENTS AND RESULTS

A. Evaluation Criteria

To test and detect our autobiographical based model of emotions, we developed a simulator a realistic crowd behaviors and the RBAS (Realistic Behavioral Agent Simulation) [36] model is proposed. In crowd situations, it is difficult to know how they will move and pass through a confined space and how the buildings and pathways should be built such that people passing through them do not have a situation of congestion and jamming and they pass easily without injuries or problems. It is cumbersome and dangerous

to bring a large number of people along and to perform the experiments; therefore it is feasible to make the realistic simulation and making the model and then collecting a real user data based on the results from that model. The crowd of people is a complex design system. To handle the complexity, one way is to model the situation comprised of individual people into simulations and making the results based on them. The algorithm is the modified form of social force model SFM [31] that simulates the agents as ellipses with particular sizes. The agents have sense of the environment and plan their own paths to avoid collisions. The shaking and repelling effect in agents have been reduced using the body contact and sliding forces. The interactions among agents have been modeled based on personal reaction space. These modifications in the SFM model make the model more realistic as every parameter has been modeled from the results of psychology and video tracking. The crowd simulation model has been developed on C++. The Figure 4 shows the instance of simulator developed. The simulation model allows creating different scenarios and range of agents. The variable numbers of agents enter into the simulator at random positions and time. The agents have random birth rate (time to come in simulation,) which is modeled by the Poisson distribution. The time at which the particles enter is also randomly timed based on normal distribution. Table 1 shows the parameters that have been used for generate scenarios and testing the proposed algorithm. Such kind of diverse variations allowed us to overcome the bottleneck of platform issues, as simulations environments are different every time due to variable parameters.

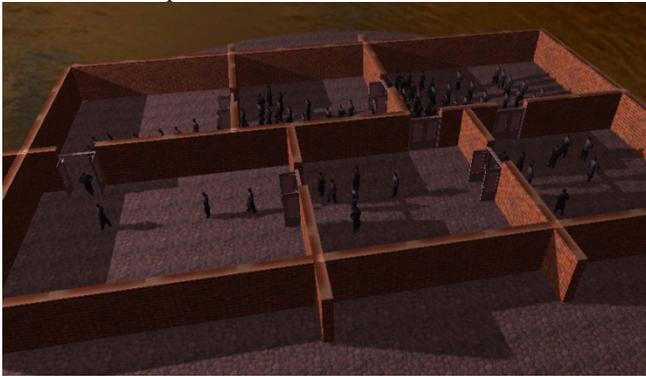


Figure 4. Visualization of developed Crowd Simulator

Parameter	Data
Duration of training set acquisition	3000 min
Number of Proto super-states observed	115610
Number of proto events observed	18251
Number of core super-states observed	112871
Number of core events observed	15478
T_{max}^-	15 sec.
T_{max}^+	40 sec.
Negative emotions events	5630
Positive emotion events	8400

Table 1. Main parameters and data during training phase

B. Experiments and Results

Different scenarios have been produced in crowd simulator at different time instants with different situations. The herding, turbulent, stop-and-go waves and normal behavior have been produced. The several aspects of the proposed framework have been evaluated. For testing purposes, 400 events of each behavior have been used with random situations and emotion classification. The normal situation has only positive emotions, while all other situations mentioned have negative emotions. The Table 2 shows the confusion matrix of classification of emotions. The red color shows the percentage of negative emotions detected in crowd, while blue demonstrates the positive emotions detected within the crowd. The percentage of emotions is derived from the analysis of the observed trajectory after execution from the environment.

Several aspects of the proposed framework have also been investigated. To measure the effectiveness, we define two parameters z_1^i and z_2^i , which are the accumulative measure of positive and negative emotions

$$z_1^i = y_{j(t-t_\theta)}^i + p(\theta_{positive}^i | \theta_w^i) \quad (14)$$

$$z_2^i = y_{j(t-t_\theta)}^i + p(\theta_{negative}^i | \theta_w^i) \quad (15)$$

- Topological coarseness: The parameter τ for ITM has been modified to check the coarseness of partitioning of environment. The Table 2 examines the proposed method presented. The fine partitioning of ($\tau = 500$) was chosen after evaluation;
- Learning: Training of AM has been performed using different training sets from simulated data and increasing number of trajectories; $n_T = \{1000, 2000, 3000, 5000\}$. The training set size is a critical issue and complexity grows exponentially as data set size increase. However, for better classification, we used 5000 trajectories. The Table 3a shows the effect on classification based on dataset size;
- Recognition measure: The correct recognition measures of positive and negative emotions in the crowd are measured and are demonstrated in the Table 3b and 3c.

	Normal	Herding	Turbulent	Stop -and-go
Normal	82.3 %	1.5 %	2.4 %	2.1 %
Herding	7.9 %	81.6 %	5.0 %	5.7 %
Turbulent	2.6 %	5.3 %	78.8 %	12.6 %
Stop -and-go waves	7.1 %	11.6 %	13.8 %	79.6 %
Total	100 %	100 %	100 %	100 %

Table 2. Confusion Matrix of classification with $\tau = 500$, $n_T = 3500$ with z_1 and z_2

	z_1	z_2
$\tau = 500$	89.2 %	86.3 %
$\tau = 1000$	78.1 %	75.9 %

(a)

z_1	500	2000	3500
Normal	91.3 %	87.2 %	80.3 %
Herding	10.3 %	15.6 %	20.1 %
Turbulent	12.3 %	14.6 %	19.8 %
Stop-and-go waves	15.8 %	25.2 %	26.1 %

(b)

z_2	500	2000	3500
Normal	12.3 %	9.3 %	8.1 %
Herding	85.6 %	82.4 %	80.5 %
Turbulent	90.2 %	88.7 %	86.2 %
Stop-and-go waves	89.4 %	78.2 %	70.8 %

(c)

Table 3. Evaluation of proposed model algorithm (a) Topological coarseness (b) Learning of AM evaluation (c) Recognition evaluation

The evaluation of the proposed algorithm shows the positive and negative emotions have been classified in the given context of an event. The parameters z_1 and z_2 demonstrate the evaluations of emotions in different kinds of behaviors. During herding behavior the accuracy drops due to the problem of classification of immobile crowd. As static crowd situation arises, the trajectory accumulations delay the classification for at least 60 frames to detect the herding behavior and then start classifying them into positive and negative trajectories.

IV. CONCLUSION

A causal event based emotion detection and classification technique has been proposed. Using the dynamic Bayesian networks, the bio-inspired autobiographical memory is generated and learned for emotion detection and classification in crowd has been demonstrated. The proposed model is the first one in the literature to present an algorithm on detection of emotions of the crowd and the results have shown that the model is capable of detecting the emotions of the people. The system is quite useful in surveillance environments to predict the crowd situation before it happens. The future work includes the collection of datasets with the help of experts and making a generalized context for all kinds of situations.

ACKNOWLEDGMENT

We are especially thankful to experts for their feedback regarding understanding situations in crowd. People from police and community management were quite helpful in this regard especially Mr. Van de Laar whose experience as a police officer helped us to understand the situations. This work was supported in part by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the EACEA Agency of the European Commission under EMJD ICE FPA n 2010-0012

REFERENCES

- [1] Bower, Gordon H., and Paul R. Cohen. "Emotional influences in memory and thinking: Data and theory." *Affect and cognition* 1 (1982).
- [2] Lazarus, Richard S. *Emotion and adaptation*. New York: Oxford University Press, 1991.
- [3] Remagnino, Paolo, and Gian Luca Foresti. "Ambient intelligence: A new multidisciplinary paradigm." *Systems, Man and Cybernetics, Part A: Systems and Humans*, IEEE Transactions on 35.1 (2005): 1-6.
- [4] Ekman, Paul. "An argument for basic emotions." *Cognition & Emotion* 6.3-4 (1992): 169-200.
- [5] Ekman, Paul. "Expression and the nature of emotion." *Approaches to emotion* 3 (1984): 19-344.
- [6] Ekman, Paul, and Wallace V. Friesen. "Facial action coding system: A technique for the measurement of facial movement. Palo Alto." CA: Consulting Psychologists Press. Ellsworth, PC, & Smith, CA (1988). From appraisal to emotion: Differences among unpleasant feelings. *Motivation and Emotion* 12 (1978): 271-302. G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-64.
- [7] Juslin, Patrik N., and Klaus R. Scherer. "Vocal expression of affect." *The new handbook of methods in nonverbal behavior research* (2005): 65-135.
- [8] Scherer, Klaus R. "Vocal communication of emotion: A review of research paradigms." *Speech communication* 40.1 (2003): 227-256.
- [9] Litman, Diane J., and Kate Forbes-Riley. "Predicting student emotions in computer-human tutoring dialogues." *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004. W.-K.
- [10] Fernandez, Raul, and Rosalind W. Picard. "Modeling drivers' speech under stress." *Speech Communication* 40.1 (2003): 145-159.
- [11] D'Mello, Sidney K., et al. "Automatic detection of learner's affect from conversational cues." *User Modeling and User-Adapted Interaction* 18.1-2 (2008): 45-80.
- [12] Mota, Selene, and Rosalind W. Picard. "Automated posture analysis for detecting learner's interest level." *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on*. Vol. 5. IEEE, 2003.
- [13] Barakova, Emilia I., and Tino Lourens. "Expressing and interpreting emotional movements in social games with robots." *Personal and ubiquitous computing* 14.5 (2010): 457-467.
- [14] Lourens, Tino, Roos Van Berkel, and Emilia Barakova. "Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis." *Robotics and Autonomous Systems* 58.12 (2010): 1256-1265.
- [15] Villon, Olivier, and Christine Lisetti. "A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors." *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*. IEEE, 2006.
- [16] Panksepp, Jaak. *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press, 1998.
- [17] Panksepp, Jaak. "Emotions as natural kinds within the mammalian brain." *Handbook of emotions* 2 (2000): 137-156.
- [18] Kapoor, Ashish, and Rosalind W. Picard. "Multimodal affect recognition in learning environments." *Proceedings of the 13th annual ACM international conference on Multimedia*. ACM, 2005.
- [19] Bänziger, Tanja, Didier Grandjean, and Klaus R. Scherer. "Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT)." *Emotion* 9.5 (2009): 691.
- [20] Le Bon, Gustave. *The crowd: A study of the popular mind*. Macmillan, 1897.
- [21] Damasio, Antonio R. *The feeling of what happens: Body and emotion in the making of consciousness*. Download iTunes eBook, 1999.
- [22] Kohonen, Teuvo. *Self-organizing maps*. Vol. 30. Springer, 2001.
- [23] Ortony, Andrew. *The cognitive structure of emotions*. Cambridge university press, 1990.
- [24] Russell, James A., Maria Lewicka, and Toomas Niit. "A cross-cultural study of a circumplex model of affect." *Journal of personality and social psychology* 57.5 (1989): 848.
- [25] Martinetz, Thomas, and Klaus Schulten. "Topology representing networks." *Neural Networks* 7.3 (1994): 507-522.

- [26] Jockusch, Jan, and Helge Ritter. "An instantaneous topological mapping model for correlated stimuli." *Neural Networks*, 1999. IJCNN'99. International Joint Conference on. Vol. 1. IEEE, 1999.
- [27] Chiappino, Simone, et al. "Event based switched dynamic bayesian networks for autonomous cognitive crowd monitoring." *Wide Area Surveillance*. Springer Berlin Heidelberg, 2014. 93-122.
- [28] Dore, Alessio, et al. "A bio-inspired system model for interactive surveillance applications." *Journal of Ambient Intelligence and Smart Environments* 3.2 (2011): 147-163.
- [29] Dore, Alessio, and Carlo S. Regazzoni. "Interaction Analysis with a Bayesian Trajectory Model." *IEEE Intelligent Systems* 25.3 (2010): 32-40.
- [30] Dore, Alessio, Andrea Fabio Cattoni, and Carlo S. Regazzoni. "Interaction modeling and prediction in smart spaces: a bio-inspired approach based on autobiographical memory." *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 40.6 (2010): 1191-1205.
- [31] Helbing, Dirk, et al. "Simulation of pedestrian crowds in normal and evacuation situations." *Pedestrian and evacuation dynamics* 21 (2002): 21-58.
- [32] Hughes, Roger L. "The flow of human crowds." *Annual review of fluid mechanics* 35.1 (2003): 169-182.
- [33] Aw, A., and Michel Rascle. "Resurrection of "second order" models of traffic flow." *SIAM journal on applied mathematics* 60.3 (2000): 916-938.
- [34] Figueiredo, Mario A. T., and Anil K. Jain. "Unsupervised learning of finite mixture models." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.3 (2002): 381-396.
- [35] Beran, Rudolf. "Minimum Hellinger distance estimates for parametric models." *The Annals of Statistics* (1977): 445-463.
- [36] Baig, Mirza Waqar, Emilia Barakova, Carlo S. Regazzoni, and Matthias Rauterberg. "Realistic Modeling of Agents in Crowd Simulations." *Intelligent Systems Modelling & Simulation (ISMS), 2014 5th International Conference on. IEEE, 2014.* 507-512.

Baig M.W., Barakova E.I., Marcenaro L., Rauterberg M., Regazzoni C.S. (2014). Bio-inspired probabilistic model for crowd emotion detection. In: *Proceedings International Joint Conference on Neural Networks - IJCNN 2014* (pp. 3966-3973), IEEE Computer Society's Conference Publishing Services.